

First-principles interatomic potentials for ten elemental metals via compressed sensing

Atsuto Seko,^{1,2,*} Akira Takahashi,¹ and Isao Tanaka^{1,2,3}

¹*Department of Materials Science and Engineering, Kyoto University, Kyoto 606-8501, Japan*

²*Center for Elements Strategy Initiative for Structure Materials (ESISM), Kyoto University, Kyoto 606-8501, Japan*

³*Nanostructures Research Laboratory, Japan Fine Ceramics Center, Nagoya 456-8587, Japan*

(Dated: May 18, 2015)

Interatomic potentials have been widely used in atomistic simulations such as molecular dynamics. Recently, frameworks to construct accurate interatomic potentials that combine a systematic set of density functional theory (DFT) calculations with machine learning techniques have been proposed. One of these methods is to use compressed sensing to derive a sparse representation for the interatomic potential. This facilitates the control of the accuracy of interatomic potentials. In this study, we demonstrate the applicability of compressed sensing to deriving the interatomic potential of ten elemental metals, namely, Ag, Al, Au, Ca, Cu, Ga, In, K, Li and Zn. For each elemental metal, the interatomic potential is obtained from DFT calculations using elastic net regression. The interatomic potentials are found to have prediction errors of less than 3.5 meV/atom, 0.03 eV/Å and 0.15 GPa for the energy, force and the stress tensor, respectively, which enable the accurate prediction of physical properties such as lattice constants and the phonon dispersion relationship.

PACS numbers: 71.15.Pd, 31.50.Bc, 34.20.-b, 65.40.-b

I. INTRODUCTION

Molecular dynamics (MD) has been a popular tool for modeling a collection of interacting atoms within classical mechanics[1]. The relationship between energy and atomic coordinates, namely, the potential energy surface (PES), plays a key role in MD simulations since the PES determines the forces acting on atoms that originate from atomic interactions and therefore the motion of atoms. As an alternative to first principles MD calculations, which provide the most accurate PES[2], frameworks to estimate a reliable PES based on the combination of systematic density functional theory (DFT) calculations and machine learning techniques have recently been proposed, which are applicable to periodic systems[3, 4]. The starting point of these methods is that a DFT calculation is performed for at least 10^3 different atomic configurations. A PES is then constructed from the DFT training data set using regression techniques to estimate the relationship between predictor and observation variables. Its accuracy is known to be much better than that of conventional interatomic potentials owing to the flexibility of the method. The flexibility also makes it possible to construct the PES for a wide range of materials using the same method.

To estimate the PES from a data set of the energy for many atomic configurations, a variety of methods can be applied. For applications to molecules and clusters, spline methods[5, 6], interpolating moving least-squares methods[7, 8], modified Shepard interpolation and other interpolation techniques[9–11], artificial neural networks[12–29] and the reproducing kernel Hilbert space method[30, 31] have been used. However, only a

few nonlinear regression techniques such as artificial neural networks[3, 32–37] and Gaussian process regression[4] have been adopted to estimate the PES for solids on account of the complex relationship between the energy and crystal structure in solids. Thus, applications to solids have been limited to a small number of metallic[32, 34], covalent[3, 4, 35] and ionic materials[33].

In these methods, the PES is generally estimated by transforming atomic positions into some descriptors. This plays an essential role in constructing a PES that satisfies several invariances, such as translational and rotational invariance, and is transferable to structures composed of a different number of atoms from those in the training data. Obviously, the accuracy of the PES strongly depends on the selection of the descriptors. So far, several descriptors for expressing atomic coordinates have been proposed[38–43], although only some of them have succeeded in obtaining accurate PESs. Descriptors are preferably chosen without a priori knowledge of the energetics of the target material. Therefore, it is desirable to establish a method that enables automatic optimization of the descriptors for constructing a PES.

The use of least absolute shrinkage and selection operator (LASSO) regression[44, 45] is a promising means of enabling the automatic selection of descriptors. We previously introduced a simple scheme to estimate the PES using LASSO regression that was based on a set of simple and systematic basis functions and a linear relationship between the energy and basis functions[46], which was applied to the elemental metals of Na and Mg. It was found that a sparse representation for the PES with a small number of basis functions was efficiently derived from relatively a large number of systematic candidate basis functions depending only on the distances between atoms. We also found that the energy can be expressed by a linear relationship with the basis functions. As a result, sparse PESs with predic-

* seko@cms.mtl.kyoto-u.ac.jp

tion errors of 1.3 and 0.9 meV/atom were obtained for Na and Mg, respectively. In addition to our application of LASSO regression for PES construction, it has recently been used to obtain sparse representations for alloy thermodynamics[47], interatomic force constants[48] and the PES for a molecule[49].

The scheme to construct the PES on the basis of LASSO regression with a linear relationship for energy has a number of advantages: 1) Accuracy can be controlled in a transparent manner. 2) A well-optimized sparse representation for the PES is obtained, which can accelerate and increase the accuracy of atomistic simulations while decreasing the computational costs. 3) Information on the forces acting on atoms and stress tensors can be included in the training data in a straightforward manner. 4) Regression coefficients are generally determined quickly using a standard least-squares technique. 5) The number of regression coefficients does not explicitly depend on the size of the input data set.

In this study, we apply this scheme to construct PESs for ten elemental metals, namely, Ag, Au, Al, Ca, Cu, Li, K, Ga, In and Zn. Here, we use elastic net regression, which is a generalization of LASSO regression. This paper is organized as follows. Sec. II presents the methodology including the linear expression for the total energy, the systematic basis functions and the regression techniques. Linear expressions for the forces acting on atoms and the stress tensors are also derived from the expression for the total energy. In Sec. III, the procedure for optimizing the input factors used to estimate the PES is described. In Sec. IV, the application of elastic net regression to the ten elemental metals is demonstrated. Finally, we give a conclusion in Sec. V.

II. METHODOLOGY

A. Linear expressions for total energy, forces acting on atoms and stress tensor

To model the relationship between the total energy and the crystal structure, we adopt the linear expression for the total energy proposed in Ref. 46. Figure 1 shows the linear model for the total energy, which is based on the widely accepted idea that the total energy of a structure is equal to the sum of its atomic energies[3, 4]. Introducing a basis expansion derived from a set of other atomic positions, this model assumes a linear relationship between the energy of atom j in structure i , $E^{(i,j)}$, and the set of basis functions for atom j in structure i , $\{b_n^{(i,j)}\}$. The linear relationship between the atomic energy and M given basis functions is expressed as

$$E^{(i,j)} = \sum_{n=0}^M w_n b_n^{(i,j)}, \quad (1)$$

where w_n and $b_n^{(i,j)}$ denote the expansion coefficients and basis functions for atom j of structure i , respectively, and

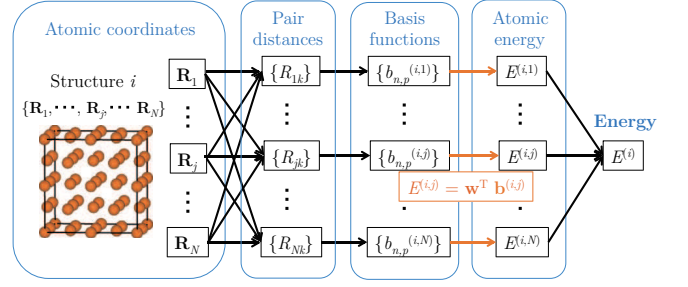


FIG. 1. Relationship between total energy and crystal structure.

$b_0^{(i,j)} = 1$. By applying the same expansion coefficients to identical atomic species, the total energy of structure i composed of $N^{(i)}$ atoms is derived as

$$\begin{aligned} E^{(i)} &= \sum_j E^{(i,j)} \\ &= \sum_n w_n \left[\sum_j b_n^{(i,j)} \right] \\ &= \sum_n w_n x_n^{(i)}, \end{aligned} \quad (2)$$

where $x_n^{(i)}$ satisfies

$$x_n^{(i)} = \sum_j b_n^{(i,j)}. \quad (3)$$

Consequently, the total energy of structure i is expressed as a linear relationship with the sum of the basis functions for all atoms in structure i .

The forces acting on atoms and the stress tensor can be given by linear equations as well as the total energy (Appendix A for details). α th component of the force acting on atom l and the virial stress tensor $\sigma_{\alpha\beta}$ of structure i are expressed as

$$F_{l,\alpha}^{(i)} = \sum_n w_n x_{\text{force},n}^{(i,l,\alpha)} \quad (4)$$

and

$$\sigma_{\alpha\beta}^{(i)} = \sum_n w_n x_{\text{stress},n}^{(i,\alpha,\beta)} \quad (5)$$

respectively, where $x_{\text{force},n}^{(i,l,\alpha)}$ and $x_{\text{stress},n}^{(i,\alpha,\beta)}$ can be derived from the derivative of the basis functions with respect to the atomic coordinates as will be shown later.

B. Estimation of regression coefficients

The expansion coefficients $\mathbf{w} = [w_0, w_1, \dots, w_M]^T$ characterizing the energetics of a system can be estimated

by regression, which is a machine learning method for estimating the relationship between the predictor and observation variables using a training data set. Regarding the training data, the energy, the forces acting on atoms and the stress tensor computed by DFT calculation can all be used as observations in the regression process, since all of them can be expressed by linear equations with the same expansion coefficients. When considering only the energy as observations, the predictor matrix \mathbf{X} and observation vector \mathbf{y} correspond to $\mathbf{X}_{\text{energy}}$ and $\mathbf{y}_{\text{energy}}$, which are composed of $x_n^{(i)}$ and the DFT energies for the structures in the training data, respectively, that is,

$$\mathbf{X} = \mathbf{X}_{\text{energy}}, \quad \mathbf{y} = \mathbf{y}_{\text{energy}}. \quad (6)$$

When using the energy, forces and stress tensor as observations, \mathbf{X} and \mathbf{y} are written as

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_{\text{energy}} \\ \mathbf{X}_{\text{force}} \\ \mathbf{X}_{\text{stress}} \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} \mathbf{y}_{\text{energy}} \\ \mathbf{y}_{\text{force}} \\ \mathbf{y}_{\text{stress}} \end{bmatrix}, \quad (7)$$

where $\mathbf{X}_{\text{force}}$ and $\mathbf{X}_{\text{stress}}$ are composed of $x_{\text{force},n}^{(i,l,\alpha)}$ and $x_{\text{stress},n}^{(i,\alpha,\beta)}$ for the structures in the training data, respectively. $\mathbf{y}_{\text{force}}$ and $\mathbf{y}_{\text{stress}}$ are composed of the forces and stress tensor computed by the DFT calculation for the structures in the training data, respectively.

A simple procedure to estimate the expansion coefficients is to use linear ridge regression. This is a shrinkage method and shrinks the regression coefficients by imposing a penalty. The ridge coefficients minimize the penalized residual sum of squares expressed as

$$L(\mathbf{w}) = \|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{w}\|_2^2, \quad (8)$$

where λ controls the magnitude of the penalty. This is referred to as L_2 regularization. The solution is easily obtained only in terms of matrix operations as $\mathbf{w} = (\mathbf{X}^\top \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^\top \mathbf{y}$, where \mathbf{I} denotes the unit matrix. Therefore, the regression coefficients can be easily estimated while avoiding the well-known multicollinearity problem occurring in the ordinary least-squares method.

Although linear ridge regression is useful for obtaining a PES from a given basis set, a basis set appropriate for the system of interest is generally unknown. Moreover, a PES with a small number of basis functions is desirable to decrease the computational cost in atomistic simulations. Therefore, we use elastic net regression[45, 50] in combination with the preparation of a considerable number of basis functions. Elastic net regression is a generalization of LASSO regression[44, 45] and combines the L_1 and L_2 penalties. Elastic net regression enables us not only to provide a solution for linear regression but also to obtain a sparse representation with a small number of nonzero regression coefficients. The solution is obtained by minimizing the function

$$L(\mathbf{w}) = \|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2^2 + \alpha \lambda \|\mathbf{w}\|_1 + \frac{(1-\alpha)}{2} \lambda \|\mathbf{w}\|_2^2, \quad (9)$$

where the parameter α determines the mixing of the penalties. When $\alpha = 1$, the minimization function corresponds to that of the LASSO. The accuracy of the solution can be controlled simply by adjusting the values of λ and α for a given training data set.

The use of elastic net regression allows us to avoid several limitations of the LASSO. For example, if there is a group of highly correlated predictor variables, the LASSO tends to select only one variable from the group. Also, in the case of high-dimensional predictor variables with few observations, the LASSO selects at most the same number of predictor variables as the number of observations before the solution saturates.

Note that the units for the energy, forces and stress tensor are different, hence care is required in the selection of the units. The units act as weights in the regression. Here, we used the units of eV/supercell, eV/Å and GPa for the energy, forces and stress tensor, respectively, when considering all of them as observations. When considering only the energy as observations, the unit of eV/atom can also be used. In this study, regression coefficients were estimated using the standardized training data.

C. Basis functions

In this study, the following simple form of the basis functions is newly used as the linear expression for the energy. The p th power of the n th element for atom j of structure i , $b_{n,p}^{(i,j)}$, is written as

$$b_{n,p}^{(i,j)} = \left[\sum_k f_n(R_{jk}^{(i)}) \cdot f_c(R_{jk}^{(i)}) \right]^p, \quad (10)$$

where p is a positive integer and $R_{jk}^{(i)}$ denotes the distance between atoms j and k of structure i . The sum is taken over all atoms within a cutoff radius R_c from atom j . $f_n(R_{jk}^{(i)})$ and $f_c(R_{jk}^{(i)})$ are an analytical pairwise function and a smooth pairwise cutoff function that is zero at a distance greater than R_c , respectively. Since the product of f_n and f_c is pairwise, an exponential form of the sum of the pairwise functions is introduced to take many-body effects into account. Although pairwise functions are adopted here for f_n , other types of basis functions such as angular basis functions can be used in principle.

Taking the sum of the basis functions for all atoms, $x_{n,p}^{(i)}$ is obtained as

$$x_{n,p}^{(i)} = \sum_j \left[\sum_k f_n(R_{jk}^{(i)}) \cdot f_c(R_{jk}^{(i)}) \right]^p. \quad (11)$$

Using a combination of the linear model and this form of the basis functions, the expression for the total energy is invariant to the translation, rotation and exchange of atoms. In addition, it can be used to input crystal structures with a different number of atoms from the structures in the training data.

$x_{\text{force},n,p}^{(i,l,\alpha)}$ and $x_{\text{stress},n,p}^{(i,\alpha,\beta)}$ can also be derived as

$$x_{\text{force},n,p}^{(i,l,\alpha)} = -p \sum_j b_{n,p-1}^{(i,j)} \frac{\partial b_{n,1}^{(i,j)}}{\partial R_{l,\alpha}^{(i)}} \quad (12)$$

and

$$x_{\text{stress},n,p}^{(i,\alpha,\beta)} = -\frac{p}{V} \sum_l R_{l,\alpha}^{(i)} \sum_j b_{n,p-1}^{(i,j)} \frac{\partial b_{n,1}^{(i,j)}}{\partial R_{l,\beta}^{(i)}}. \quad (13)$$

respectively (Appendix A for details). The derivative of the basis functions with respect to the α th component of the atomic position is written as

$$\frac{\partial b_{n,1}^{(j)}}{\partial R_{l,\alpha}} = \sum_k [f'_n(R_{jk})f_c(R_{jk}) + f_n(R_{jk})f'_c(R_{jk})] \frac{\partial R_{jk}}{\partial R_{l,\alpha}}, \quad (14)$$

where the structure index i is omitted.

For the pairwise analytical function f_n , we introduce Gaussian, cosine, Bessel, Neumann, modified Morlet wavelet (MMW), Slater-type orbital (STO) and Gaussian-type orbital (GTO) functions. Table I shows the different function forms of f_n and their derivatives with respect to the distance f'_n used in this study. The derivatives can be seen in the expressions for the forces acting on atoms and the stress tensor. For the cosine and MMW types, function forms with a single internal parameter are introduced, while for the Gaussian, STO and GTO types, function forms with two internal parameters are used. Using a number of functions with different internal parameters for each type of function, a systematic set of basis functions used to select the basis in elastic net regression is obtained. For the cutoff function, we adopt the cosine-based cutoff function used in Ref. 3, expressed as

$$f_c(R) = \begin{cases} \frac{1}{2} \left[\cos\left(\pi \frac{R}{R_c}\right) + 1 \right] & (R \leq R_c) \\ 0 & (R > R_c) \end{cases}. \quad (15)$$

III. OPTIMIZATION OF INPUT FACTORS

The accuracy of the elastic net PES mainly depends on 1) the cutoff radius R_c , 2) the size of the training data, 3) the variety of structures included in the training data, 4) the observation properties used for regression, 5) the candidate basis functions and 6) the parameters α and λ in the minimization function of the elastic net regression. However, it is difficult to optimize all of these input factors simultaneously. We therefore optimize them in a stepwise manner.

A. DFT data set

To begin with, training and test data sets were generated from systematic DFT calculations. The test data set

was used to examine the predictive power for structures that were not included in the training data set. We generated 2700 and 300 configurations for the training and test data sets, respectively, for each elemental metal. They include structures made by isotropic expansions, random expansions and random distortions of ideal face-centered-cubic (fcc), body-centered-cubic (bcc), hexagonal-closed-packed (hcp), simple cubic (sc), ω and β -tin structures, in which the atomic positions and lattice constants were fully optimized. Random structures were generated using Gaussian random numbers with ten different values for the variance. These configurations were made using supercells constructed by the $2 \times 2 \times 2$, $3 \times 3 \times 3$, $3 \times 3 \times 3$, $4 \times 4 \times 4$, $3 \times 3 \times 3$ and $2 \times 2 \times 2$ expansions of the conventional unit cells for fcc, bcc, hcp, sc, ω and β -tin structures, which are composed of 32, 54, 54, 64, 81 and 32 atoms, respectively.

For a total of 3000 configurations for each elemental metal, DFT calculations were performed using the plane-wave basis projector augmented wave (PAW) method[51, 52] within the Perdew-Burke-Ernzerhof exchange-correlation functional[53] as implemented in the VASP code[54, 55]. The cutoff energy was set to 400 eV. The total energies converged to less than 10^{-3} meV/supercell. For only the ideal structures, the atomic positions and lattice constants were optimized until the residual forces became less than 10^{-3} eV/Å.

B. Cutoff radius

The optimal cutoff radius was determined for each elemental metal using linear ridge regression with a given set of 180 basis functions consisting of 18 Bessel, 18 Neumann, 60 cosine and 84 Gaussian functions. We searched for a cutoff radius giving a low prediction error by constructing PESs using several cutoff radii ranging from 5 to 10 Å with intervals of 0.5 Å. The ridge penalty term was set to $\lambda = 10^{-4}$. To estimate the prediction error of each PES, we calculated the root-mean-square error (RMSE) between the observation property predicted by the DFT calculation and that predicted by the PES for the test data. Then, the convergence of the RMSE was examined. Here, the energy, the force and the stress tensor were used as the observation properties. In addition, the cutoff radius also plays an essential role in expressing the energy of structures with large volumes. However, the contribution of such structures to the RMSE was minor in our data sets. Therefore, we determined the cutoff radii using the convergence of the energy-volume curve in addition to the convergence of the RMSE.

Table II shows the optimized cutoff radii. Hereafter, these optimized values will be used. Table II also shows the RMSEs of the energy, force and stress tensor. We obtained PESs with the RMSEs for the energy ranging from 0.5 to 4.0 meV/atom. In other words, PESs with high accuracy were obtained even using a given set of basis functions and linear ridge regression. The given set

TABLE I. Adopted analytical pairwise functions for $f_n(R)$ and their derivatives $f'_n(R)$. The internal parameters a and b are always positive except for parameter a in the STO and GTO functions. For the Bessel and Neumann functions, n is a positive integer.

	$f_n(R)$	$f'_n(R)$
Bessel	$J_n(R)$	$-f_1(R)$ ($n=0$), $[f_{n-1}(R) - f_{n+1}(R)]/2$ ($n \geq 1$)
Neumann	$Y_n(R)$	$-f_1(R)$ ($n=0$), $[f_{n-1}(R) - f_{n+1}(R)]/2$ ($n \geq 1$)
Cosine	$\cos(aR)$	$-a \sin(aR)$
Modified Morlet wavelet (MMW)	$\cos(aR)/\cosh(R)$	$-a \sin(aR)/\cosh(R) - f_n(R) \tanh(R)$
Gaussian	$\exp[-a(R-b)^2]$	$-2a(R-b)f_n(R)$
Slater-type orbital (STO)	$R^a \exp(-bR)$	$(a-bR)R^{a-1} \exp(-bR)$
Gaussian-type orbital (GTO)	$R^a \exp(-bR^2)$	$(a-2bR^2)R^{a-1} \exp(-bR^2)$

TABLE II. Optimal cutoff radius R_c . PESs are constructed by linear ridge regression using a given basis set composed of 180 basis functions including 18 Bessel, 18 Neumann, 60 cosine and 84 Gaussian functions.

Linear ridge regression ($N_{\text{basis}} = 180$)				
	R_c (Å)	RMSE (energy) (meV/atom)	RMSE (force) (eV/Å)	RMSE (stress) (GPa)
Ag	7.5	2.4	0.012	0.08
Al	8.0	4.0	0.020	0.15
Au	6.0	3.0	0.030	0.16
Ca	9.5	1.2	0.011	0.03
Cu	7.5	2.6	0.018	0.10
Ga	10.0	1.9	0.019	0.11
In	10.0	1.9	0.017	0.07
K	10.0	0.5	0.001	0.00
Li	8.5	0.5	0.005	0.02
Zn	10.0	3.0	0.021	0.22

of basis functions may be acceptable for expressing the energetics of the ten different elemental metals. However, this may not be the case with other systems.

C. Observation property

The dependence of the PES accuracy on the observation property of the training data was examined. We compared two training sets of observation properties used for regression. One training set was composed only of the energy, and the other set is composed of the energy, force and stress tensor. The comparison was carried out using the above 180 basis functions, the optimized cutoff radius and linear ridge regression. When using only the energy as the observation property, the RMSE was small for the energy but large for the force and the stress tensor. Including the force and the stress tensor to the observation properties resulted in improved prediction for the force and the stress tensor at the expense of the predictive power for the energy. To ensure accuracy for both the

TABLE III. Optimized candidate basis set used in elastic net regression. Min. and Max. stand for the minimum and maximum values of the arithmetic sequence of the internal parameter, respectively. N_{seq} denotes the number of components of the sequence.

Basis type	Number of basis functions	Internal parameter			
			Min.	Max.	N_{seq}
Bessel	18	n	0	5	6
Neumann	18	n	0	5	6
Cosine	300	a	0.1	10.0	100
MMW	300	a	0.1	10.0	100
Gaussian	1200	a	0.1	2.0	20
		b	0.0	5.0	20
STO	1500	a	-2	2	5
		b	0.1	10.0	100
GTO	1500	a	-2	2	5
		b	0.1	10.0	100
Total	4836				

force and the stress tensor, which is essential for calculations of phonon dispersions and structure optimization, we hereafter use the energy, force and stress tensor as the observation properties unless otherwise specified.

D. Candidate basis set

Even using a combination of linear ridge regression and the above 180 basis functions, PESs with high accuracy were sometimes obtained. However, a PES with a smaller number of basis functions is generally preferable to accelerate the computation of the energy, forces and stress tensors. Additionally, it may be possible to find more suitable basis functions by considering other basis functions. Therefore, it is useful to select suitable basis functions from a candidate basis set including other basis functions by elastic net regression.

In general, a candidate basis set should ideally be compact for the following reasons. 1) If the elastic net uses

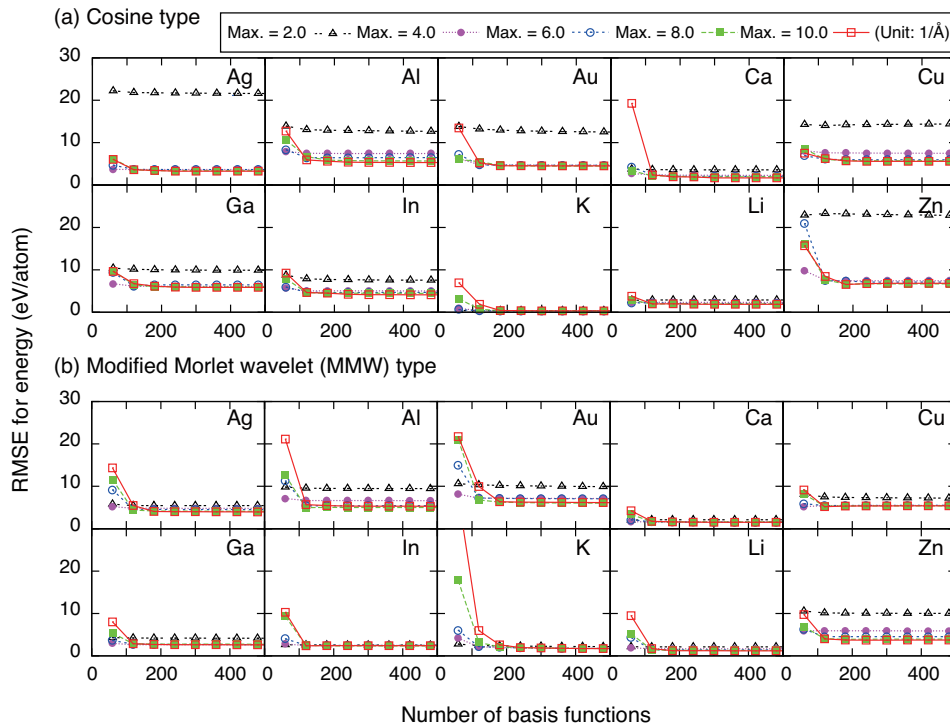


FIG. 2. Dependence of RMSE of linear ridge PES on the number of basis functions for (a) cosine and (b) MMW types. Max. stands for the maximum value of the arithmetic sequence of the internal parameter.

too many basis functions compared to the number of input observations, the selection of a good set of basis functions tends to be difficult. 2) The amount of available memory on computers can be exhausted, particularly when the forces and stress tensor are used as the observations.

A compact candidate basis set was obtained by optimization on the basis of the results of linear ridge regression with a single type of basis function. For a basis type with a single internal parameter, a trial set for the internal parameter was given by an arithmetic sequence. The sequence can be specified by the minimum and maximum values and the number of components of the sequence. For the Bessel and Neumann types, we set the minimum and interval of the arithmetic sequence to zero and one, respectively. For the cosine and MMW functions, the minimum value was taken to be 0.1. Linear ridge PESs were then constructed using many trial sets for each basis type. Here, all the basis functions with $p = 1, 2, 3$ for each f_n were considered because it has been shown that the use of $p = 1, 2, 3$ terms greatly decreases the prediction error for elemental Na and Mg[46]. The ridge penalty term was taken to be $\lambda = 10^{-4}$.

Figures 2 (a) and (b) show the convergence of the RMSE for the energy with respect to the number of basis functions for the cosine and MMW types. Unfixed parameters of the sequence were determined from the convergence of the RMSE. Table III shows the optimized number of basis functions together with the minimum

and maximum values of the sequence. We also tested polynomial forms, three types of exponential forms and Mexican hat wavelets with a single internal parameter as candidates of f_n . However, they showed a larger RMSE or unstable behavior when performing regression.

For each basis type with two internal parameters, two arithmetic sequences were given and all combinations of their components were considered. For the Gaussian type, the minimum values of the sequences for a and b were fixed to 0.1 and 0.0, respectively. The number of components of each sequence was set to 20. Therefore, each sequence was specified only by the maximum value. For the STO and GTO types, the sequence for a was fixed so that the minimum value, the interval and the maximum value were given as -2 , 1 and 2 , respectively. The minimum value of the sequence for b was set to 0.1. The sequence for b was specified by the maximum value and the number of components of the sequence. Table III shows the optimized sequences. We used all the types of basis functions shown in Table III as the candidate basis set. The total number of functions in the candidate basis set was 4836.

IV. ELASTIC NET POTENTIAL ENERGY SURFACE

PESs were then estimated by elastic net regression using the DFT observations, the candidate basis set and

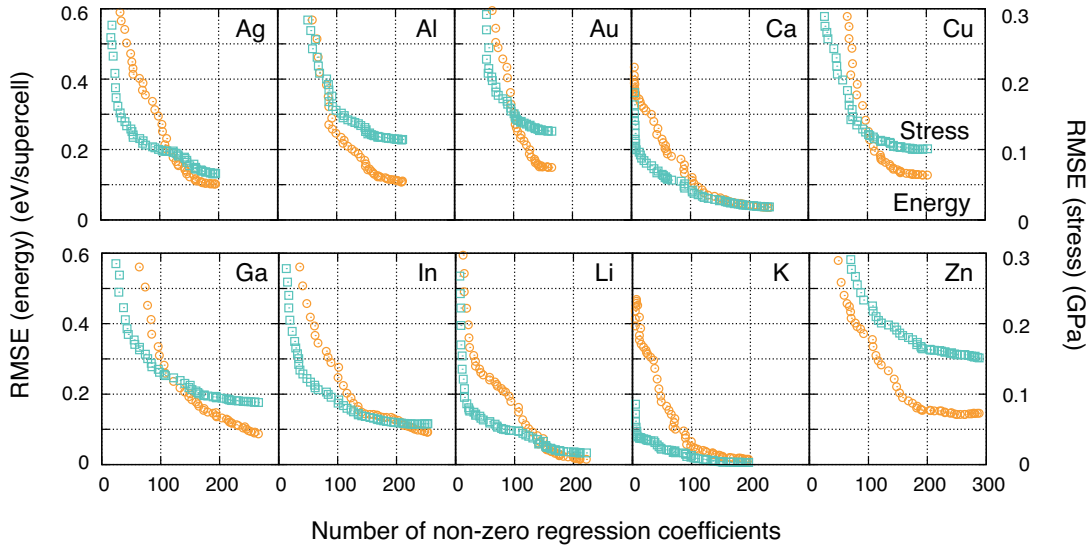


FIG. 3. Dependence of RMSEs for the energy and the stress tensor of elastic net PES ($\alpha = 1$, LASSO) on the number of the basis functions for ten elemental metals. RMSEs for the energy and the stress tensor are shown by orange open circles and blue open squares, respectively.

TABLE IV. RMSEs for the energy, the force and the stress tensor of elastic net PESs showing the minimum criterion score. Equilibrium lattice constants for the bcc and fcc structures estimated from the elastic net PES are also shown. Values in brackets were obtained directly by DFT calculation.

Element	Number of basis functions	RMSE (energy) (meV/atom)	RMSE (force) (eV/Å)	RMSE (stress) (GPa)	a (bcc) (Å)	a (fcc) (Å)
Ag	190	2.2	0.011	0.07	3.309 (3.311)	4.157 (4.160)
Al	210	3.5	0.020	0.12	3.234 (3.233)	4.039 (4.038)
Au	165	2.4	0.030	0.15	3.316 (3.309)	4.172 (4.164)
Ca	234	1.2	0.010	0.03	4.383 (4.381)	5.522 (5.519)
Cu	202	2.6	0.018	0.12	2.885 (2.887)	3.630 (3.633)
Ga	266	2.2	0.017	0.09	3.371 (3.371)	4.227 (4.228)
In	253	2.3	0.019	0.07	3.814 (3.815)	4.797 (4.797)
K	197	0.3	0.001	0.00	5.284 (5.283)	6.666 (6.662)
Li	222	0.4	0.005	0.02	3.440 (3.439)	4.329 (4.331)
Zn	288	2.9	0.016	0.15	3.130 (3.136)	3.928 (3.935)

the optimized cutoff radius. PESs were optimized by changing the parameters α and λ in the minimization function of Eqn. (9). We varied λ from 10^3 to 10^{-3} and adopted values of α of 0.6, 0.8 and 1. In elastic net regression, we used only the energy and stress tensor as the observations because the available computational memory was limited. Although the regression coefficients obtained by the elastic net were valid as they were, we reestimated them using linear ridge regression, where the energy, forces, and stress tensor were used as the observations.

As a criterion score to determine the optimal PES, the average of the RMSEs for the energy and the stress tensor was used. Here, we regarded the PES with the lowest criterion score as the optimal one. Note, however, that the

definition of the optimal PES depends on the purpose. In another situation, a PES with a smaller number of basis functions may be regarded as the optimal one when a decrease in the computational costs at the expense of slight degradation of the accuracy is desired.

Figure 3 shows the dependence of the RMSE for the energy and stress tensor on the number of basis functions when $\alpha = 1$. The number of selected basis functions tended to increase with decreasing λ . At the same time, the RMSE for the energy and stress tensor tended to decrease. Although multiple PESs with the same number of basis functions were sometimes obtained from different values of λ , only the PES with the lowest criterion score among the PESs with the same number of basis functions is shown in Fig. 3. On the other hand, the criterion score

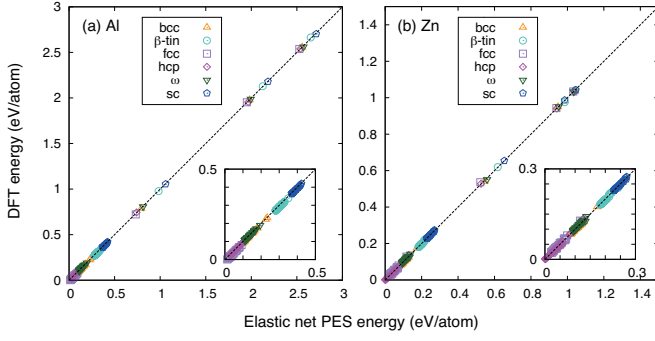


FIG. 4. Comparison of the energies predicted by the elastic net PES and DFT for (a) Al and (b) Zn, measured from the energy of the most stable structure among the bcc, fcc, hcp, sc, ω and β -tin structures.

does not change significantly with decreasing α although the number of selected basis functions increases. Therefore, we will hereafter show only the results for $\alpha = 1$.

Table IV shows the RMSEs for the energy, the force and the stress tensor of the optimal elastic net PES. We obtained PESs with the RMSE for the energy in the range of 0.3–3.5 meV/atom for the ten elemental metals using only 165–288 basis functions. The RMSEs for the force and the stress are within 0.03 eV/Å and 0.15 GPa, respectively. Compared with the RMSEs of the PESs constructed from the given basis set shown in Table II, the prediction errors are reduced for some elemental metals as a consequence of the automatic optimization of the basis set. Table IV also shows the equilibrium lattice constants of the bcc and fcc structures for the ten elemental metals predicted by the elastic net PES, together with those predicted by DFT. The PESs have equilibrium lattice constants that are in agreement with those obtained by DFT.

Figure 4 shows a comparison of the energies of test data predicted by the elastic net PES and DFT for Al and Zn, showing the largest and second largest RMSEs for the energy. As can be seen in Fig. 4, there is little difference between the DFT and elastic net PES energies regardless of the crystal structure. In addition, no dependence of the RMSE on the energy can be clearly observed despite the wide range of structures included in both the training and test data.

The applicability of the elastic net PES to the calculation of the force was also examined by comparing phonon dispersion relationships computed by the elastic net PES and DFT. The phonon dispersion relationships were calculated by the supercell approach[56] for the bcc and fcc structures with the equilibrium lattice constant. To evaluate the dynamical matrix, each symmetrically independent atomic position was displaced by 0.01 Å. The forces acting on atoms by the elastic net PES can then be analytically computed using Eqn. (4). Supercells were made by $4 \times 4 \times 4$ expansion of the conventional unit cells for both the bcc and fcc structures. The phonon calculations

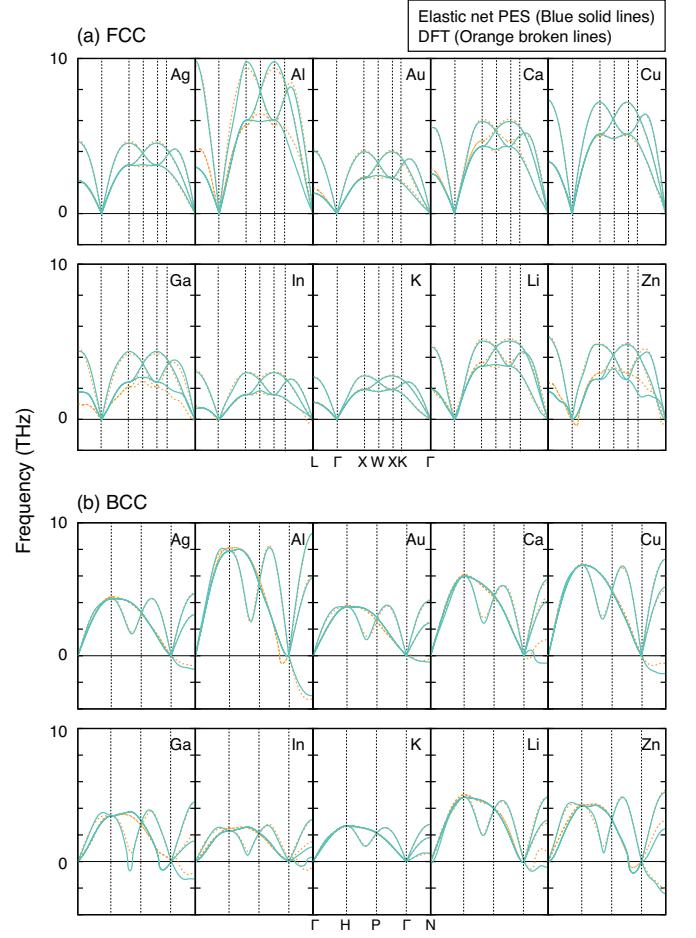


FIG. 5. Phonon dispersion relationships for ten elemental metals with (a) fcc and (b) bcc structures. Phonon dispersion curves obtained by the elastic net PES and DFT are shown by blue solid and orange broken lines, respectively. Negative values indicate imaginary modes.

were performed using the PHONOPY code[57]. Figure 5 shows the phonon dispersion relationships of the (a) bcc and (b) fcc structures for the ten elemental metals, computed by both the elastic net PES and DFT. For all elemental metals with both the bcc and fcc structures, the phonon dispersion relationships calculated by the elastic net PES are in good agreement with those calculated by DFT. This demonstrates that the elastic net PES is sufficiently accurate to perform atomistic simulations with similar accuracy to DFT calculation.

V. CONCLUSION

We have applied a method of constructing a linearized PES by elastic net regression to a wide range of elemental metals. Compared with the other approach based on systematic first-principles calculations, the elastic net interatomic potential has the following main advantages. 1)

A well-optimized sparse representation for the PES can be obtained, which increases the accuracy of atomistic simulations while decreasing the computational cost. 2) The accuracy can be easily controlled, i.e., the trade-off between the accuracy and computational cost is determined by a small number of parameters. 3) Information on the forces acting on atoms and stress tensors can be included in the training data in a straightforward manner. This ensures the reliability of the force and stress tensor calculation using constructed interatomic potentials.

As a result of applying the present method, we found that the energetics can be expressed by a linear relationship with simple basis functions depending only on distances between atoms. A sparse set of suitable basis functions for expressing the PES can also be easily extracted from 4836 basis functions by elastic net regression. As a result, we have obtained a sparse PES with prediction errors ranging from 0.3 to 3.5 meV/atom. The prediction errors for the force and the stress tensor were within 0.03 eV/Å and 0.15 GPa, respectively. Also, we compared equilibrium lattice constants and phonon dispersion relationships obtained by the elastic net PES and by DFT calculation. The former were in good agreement with the latter for all ten elemental metals considering in this study.

ACKNOWLEDGMENTS

This study was supported by a Grant-in-Aid for Scientific Research (A) and a Grant-in-Aid for Scientific Research on Innovative Areas “Nano Informatics” (grant number 25106005) from Japan Society for the Promotion of Science (JSPS). AS also acknowledges a Grant-in-Aid for Scientific Research (B) from JSPS.

Appendix A: Expressions for forces acting on atoms and stress tensor

Here, expressions for forces acting on atoms and the stress tensor are derived from the derivative of Eqn. (2) with respect to the atomic positions provided in Cartesian coordinates. Although they depend on the form of the basis functions, expressions for a linear model involving only basis functions depending only on pair distances are derived. Since the total energy of structure i has a linear relationship with the sum of the basis functions, α th component of the force acting on atom l of structure

i is expressed by

$$\begin{aligned} F_{l,\alpha}^{(i)} &= -\frac{\partial E^{(i)}}{\partial R_{l,\alpha}^{(i)}} \\ &= -\sum_{n,p} w_{n,p} \frac{\partial x_{n,p}^{(i)}}{\partial R_{l,\alpha}^{(i)}} \\ &= \sum_{n,p} w_{n,p} x_{\text{force},n,p}^{(i,l,\alpha)}, \end{aligned} \quad (\text{A1})$$

where

$$\begin{aligned} x_{\text{force},n,p}^{(i,l,\alpha)} &= -\frac{\partial x_{n,p}^{(i)}}{\partial R_{l,\alpha}^{(i)}} \\ &= -p \sum_j b_{n,p-1}^{(i,j)} \frac{\partial b_{n,1}^{(i,j)}}{\partial R_{l,\alpha}^{(i)}}. \end{aligned} \quad (\text{A2})$$

The stress tensor is generally obtained by virial stress computation. The virial stress tensor $\sigma_{\alpha\beta}$ is expressed as

$$\sigma_{\alpha\beta}^{(i)} = \frac{1}{V} \sum_l R_{l,\alpha}^{(i)} F_{l,\beta}^{(i)}, \quad (\text{A3})$$

where V denotes the volume of the cell containing $N^{(i)}$ atoms. Using the expression for the forces in Eqn. (A1), the stress tensor is derived as the following linear equation:

$$\begin{aligned} \sigma_{\alpha\beta}^{(i)} &= -\frac{1}{V} \sum_l R_{l,\alpha}^{(i)} \left[\sum_{n,p} w_{n,p} \left(p \sum_j b_{n,p-1}^{(i,j)} \frac{\partial b_{n,1}^{(i,j)}}{\partial R_{l,\beta}^{(i)}} \right) \right] \\ &= \sum_{n,p} w_{n,p} \left[-\frac{p}{V} \sum_l R_{l,\alpha}^{(i)} \sum_j b_{n,p-1}^{(i,j)} \frac{\partial b_{n,1}^{(i,j)}}{\partial R_{l,\beta}^{(i)}} \right] \\ &= \sum_{n,p} w_{n,p} x_{\text{stress},n,p}^{(i,\alpha,\beta)}, \end{aligned} \quad (\text{A4})$$

where

$$x_{\text{stress},n,p}^{(i,\alpha,\beta)} = -\frac{p}{V} \sum_l R_{l,\alpha}^{(i)} \sum_j b_{n,p-1}^{(i,j)} \frac{\partial b_{n,1}^{(i,j)}}{\partial R_{l,\beta}^{(i)}}. \quad (\text{A5})$$

By computing all the contributions from atoms within the cutoff radius using Eqn. (A4), the virial stress is obtained.

The derivative of the basis functions with respect to the α th component of the atomic position is written as

$$\begin{aligned} \frac{\partial b_{n,1}^{(j)}}{\partial R_{l,\alpha}} &= \sum_k \left[\frac{\partial f_n(R_{jk})}{\partial R_{l,\alpha}} f_c(R_{jk}) + f_n(R_{jk}) \frac{\partial f_c(R_{jk})}{\partial R_{l,\alpha}} \right] \\ &= \sum_k [f'_n(R_{jk}) f_c(R_{jk}) + f_n(R_{jk}) f'_c(R_{jk})] \frac{\partial R_{jk}}{\partial R_{l,\alpha}}, \end{aligned} \quad (\text{A6})$$

where structure index i is omitted. Three types of derivatives can be found in Eqn. (A6). The derivative of the distance with respect to the α th component of the atomic position is expressed as

$$\frac{\partial R_{jk}}{\partial R_{l,\alpha}} = \begin{cases} \frac{(\mathbf{R}_j - \mathbf{R}_k)_\alpha}{R_{jk}} & (l = j) \\ \frac{(\mathbf{R}_k - \mathbf{R}_j)_\alpha}{R_{jk}} & (l = k) \end{cases}, \quad (\text{A7})$$

where \mathbf{R}_j denotes the three-dimensional atomic position of atom j in Cartesian coordinates. The derivative of the cutoff function with respect to the distance is given by

$$f'_c(R_{jk}) = -\frac{\pi}{2R_c} \sin\left(\pi \frac{R_{jk}}{R_c}\right). \quad (\text{A8})$$

The derivative of the pairwise function f_n with respect to the distance depends on the selection of the functions, hence the expression for the derivative for each type of f_n is shown in Table I.

- [1] B. J. Alder and T. E. Wainwright, J. Chem. Phys. **31**, 459 (1959).
- [2] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).
- [3] J. Behler and M. Parrinello, Phys. Rev. Lett. **98**, 146401 (2007).
- [4] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Phys. Rev. Lett. **104**, 136403 (2010).
- [5] J. M. Bowman, J. S. Bittman, and L. B. Harding, J. Chem. Phys. **85**, 911 (1986).
- [6] S. Chapman, M. Dupuis, and S. Green, Chem. Phys. **78**, 93 (1983).
- [7] R. Dawes, D. L. Thompson, Y. Guo, A. F. Wagner, and M. Minkoff, J. Chem. Phys. **126**, 184108 (2007).
- [8] R. Dawes, D. L. Thompson, A. F. Wagner, and M. Minkoff, J. Chem. Phys. **128**, 084107 (2008).
- [9] J. Ischtwan and M. A. Collins, J. Chem. Phys. **100**, 8080 (1994).
- [10] M. J. T. Jordan, K. C. Thompson, and M. A. Collins, J. Chem. Phys. **102**, 5647 (1995).
- [11] R. Fournier and S. Orel, J. Chem. Phys. **139**, 234110 (2013).
- [12] S. Lorenz, A. Groß, and M. Scheffler, Chem. Phys. Lett. **395**, 210 (2004).
- [13] H. Gassner, M. Probst, A. Lauenstein, and K. Hermanson, J. Phys. Chem. A **102**, 4596 (1998).
- [14] F. V. Prudente, P. H. Acioli, and J. J. S. Neto, J. Chem. Phys. **109**, 8801 (1998).
- [15] F. V. Prudente and J. J. S. Neto, Chem. Phys. Lett. **287**, 585 (1998).
- [16] D. F. R. Brown, M. N. Gibbs, and D. C. Clary, J. Chem. Phys. **105**, 7597 (1996).
- [17] T. B. Blank, S. D. Brown, A. W. Calhoun, and D. J. Doren, J. Chem. Phys. **103**, 4129 (1995).
- [18] B. G. Sumpter and D. W. Noid, Chem. Phys. Lett. **192**, 455 (1992).
- [19] J. Behler, S. Lorenz, and K. Reuter, J. Chem. Phys. **127**, 014705 (2007).
- [20] J. Li, B. Jiang, and H. Guo, J. Chem. Phys. **139**, 204103 (2013).
- [21] B. Jiang and H. Guo, J. Chem. Phys. **141**, 034109 (2014).
- [22] C. M. Handley and P. L. A. Popelier, J. Phys. Chem. A **114**, 3371 (2010).
- [23] P. M. Agrawal, M. Malshe, R. Narulkar, L. M. Raff, M. Hagan, S. Bukkapatnam, and R. Komanduri, J. Phys. Chem. A **113**, 869 (2009).
- [24] H. T. T. Nguyen and H. M. Le, J. Phys. Chem. A **116**, 4629 (2012).
- [25] H. M. Le and L. M. Raff, J. Phys. Chem. A **114**, 45 (2010).
- [26] M. G. Darley, C. M. Handley, and P. L. A. Popelier, J. Chem. Theory Comput. **4**, 1435 (2008).
- [27] K.-H. Cho, K. T. No, and H. A. Scheraga, J. Mol. Struct. **641**, 77 (2002).
- [28] M. Malshe, A. Pukritayakamee, L. M. Raff, M. Hagan, S. Bukkapatnam, and R. Komanduri, J. Chem. Phys. **131**, 124127 (2009).
- [29] S. Manzhos and T. Carrington, J. Chem. Phys. **125**, 084109 (2006).
- [30] T. Hollebeek, T.-S. Ho, and H. Rabitz, Annu. Rev. Phys. Chem. **50**, 537 (1999).
- [31] T.-S. Ho and H. Rabitz, J. Chem. Phys. **119**, 6433 (2003).
- [32] H. Eshet, R. Z. Khaliullin, T. D. Kühne, J. Behler, and M. Parrinello, Phys. Rev. B **81**, 184107 (2010).
- [33] N. Artrith, T. Morawietz, and J. Behler, Phys. Rev. B **83**, 153101 (2011).
- [34] N. Artrith and J. Behler, Phys. Rev. B **85**, 045439 (2012).
- [35] G. C. Sossio, G. Miceli, S. Caravati, J. Behler, and M. Bernasconi, Phys. Rev. B **85**, 174103 (2012).
- [36] J. Behler, Phys. Chem. Chem. Phys. **13**, 17930 (2011).
- [37] J. Behler, J. Phys.: Condens. Matter **26**, 183001 (2014).
- [38] J. Behler, J. Chem. Phys. **134**, 074106 (2011).
- [39] K. J. Jose, N. Artrith, and J. Behler, J. Chem. Phys. **136**, 194111 (2012).
- [40] A. P. Bartók, R. Kondor, and G. Csányi, Phys. Rev. B **87**, 184115 (2013).
- [41] M. Rupp, A. Tkatchenko, K.-R. Müller, and O. von Lilienfeld, Phys. Rev. Lett. **108**, 058301 (2012).
- [42] K. T. Schütt, H. Glawe, F. Brockherde, A. Sanna, K. R. Müller, and E. K. U. Gross, Phys. Rev. B **89**, 205118 (2014).
- [43] O. A. von Lilienfeld, M. Rupp, and A. Knoll, arXiv preprint arXiv:1307.2918 (2013).
- [44] R. Tibshirani, J. R. Stat. Soc. B **58**, 267 (1996).
- [45] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed. (Springer, New York, 2009).
- [46] A. Seko, A. Takahashi, and I. Tanaka, Phys. Rev. B **90**, 024101 (2014).
- [47] L. J. Nelson, G. L. W. Hart, F. Zhou, and V. Ozoliņš, Phys. Rev. B **87**, 035125 (2013).
- [48] F. Zhou, W. Nielson, Y. Xia, and V. Ozoliņš, Phys. Rev. Lett. **113**, 185501 (2014).

- [49] W. Mizukami, S. Habershon, and D. P. Tew, J. Chem. Phys. **141**, 144310 (2014).
- [50] H. Zou and T. Hastie, J. R. Stat. Soc. B **67**, 301 (2005).
- [51] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).
- [52] G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).
- [53] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
- [54] G. Kresse and J. Hafner, Phys. Rev. B **47**, 558 (1993).
- [55] G. Kresse and J. Furthmüller, Phys. Rev. B **54**, 11169 (1996).
- [56] K. Parlinski, Z.-Q. Li, and Y. Kawazoe, Phys. Rev. Lett. **78**, 4063 (1997).
- [57] A. Togo, F. Oba, and I. Tanaka, Phys. Rev. B **78**, 134106 (2008).